Chapter 13
Nucleic Acid Biotechnology Techniques

SUMMARY

Section 13.1

- In order to study nucleic acids, we must have a way to separate and identify them.
- The most common separation technique is gel electrophoresis. DNA pieces separate on a gel based on size.
- DNA can be visualized on a gel via autoradiography, a procedure where the DNA has incorporated radioactive nucleotides.
- DNA can also be seen using fluorescence, often with a compound called ethidium bromide, which glows orange under UV light when bound to DNA. The disadvantage is that ethidium bromide is a strong carcinogen.
- Other dyes, such as SyBr green, that are not carcinogenic, are now being used to avoid the environmental hazards.

Section 13.2

- Restriction endonucleases are enzymes produced by bacteria that hydrolyze the phosphodiester backbone of DNA at specific sequences.
- The sequences targeted by restriction endonucleases are palindromes, meaning their sequence reads the same on both strands going in the same direction.
- Most restriction enzymes cut DNA in a way that leaves sticky ends that are very useful for recombining DNA from different sources.

Section 13.3

- Cloning refers to creating a genetically identical population.
- DNA can be combined by using restriction enzymes that create sticky ends in the DNA. This recombinant DNA will have a target DNA sequence that the experimenter is interested in.
- The target DNA sequence is carried in some type of vector, usually a bacterial plasmid or a virus.
- The target DNA sequence is inserted into a host organism and the natural doubling time of the organism is used to create many copies of the target DNA sequence.
- Those organisms that are carrying the target DNA are identified through a process called selection. Selection often involves antibiotic resistance.

Section 13.4

- Genetic engineering is the process of inserting genes of interest into specific organisms for either a medical or purely scientific benefit.
- Gene therapy is the process of inserting a missing gene into an organism.
- Bacteria are often used as the factories to produce a protein from a cloned gene. This has led to the production of human proteins such as insulin and erythropoietin.

- To produce the protein product of a gene of interest, the gene must be cloned into an expression vector, usually a plasmid with special features that allows it to be transcribed and translated in a host cell.
- In agriculture, genetic engineering is used to produce crops that are resistant to insects or have long shelf lives.

Section 13.5

- A DNA library is a collection of clones of an entire genome.
- The genome is digested with restriction enzymes and the pieces are cloned into vectors and transformed into cell lines.
- Specific radioactive probes to a sequence of interest are reacted to filters that have copies of the bacterial colonies in the library. The probe will bind to the sequence of interest and the colony's location can be seen via autoradiography.
- A cDNA library is constructed by using reverse transcriptase to make DNA from the mRNA in a cell. This cDNA is then used to construct a library similar to a genomic DNA library.

Section 13.6

- Polymerase Chain Reaction (PCR) is a sophisticated, automated technique for amplifying DNA from very small amounts of sample.
- The DNA to be amplified is mixed with specific primers, dNTPs, and a heat stabile form of DNA polymerase.
- The mixture undergoes 20-40 rounds of DNA polymerization via cycling the temperatures so that the DNA strands separate, the primers anneal, and the polymerase fills in the DNA. Each cycle doubles the target DNA.
- The technique has revolutionized forensic science as DNA can be amplified from just a few cells and then the DNA analyzed and identified
- A newer PCR technique, called quantitative PCR (qPCR) can be used to measure the original amount of DNA in a sample being amplified.

Section 13.7

- A DNA fingerprint is created by digesting DNA with restriction enzymes, separating the pieces on a gel, and then visualizing some of the pieces by using labeled probes.
- Differences in DNA patterns between different individuals are based on different base sequences of their DNA. These base sequence differences mean there will be different restriction sites leading to different length fragments.

Section 13.8

- DNA can be sequenced by using several techniques. One of the most common is called the chain termination method.
- Dideoxy nucleotides are used to terminate DNA synthesis. Multiple reactions are run with a different dideoxy nucleotide in each reaction mix.
- The reactions produce a series of DNA fragments of different length that can be run on a gel and the sequence determined by tracking the different length fragments in the lanes with the 4 different dideoxy nucleotides

Section 13.9

- As more DNA sequences become available, it becomes possible to compare those sequences. Of particular interest is any pattern that may emerge from genes that encode proteins with similar functions.
- Important medical applications are emerging, and new methods are making it possible to analyze large quantities of data. Complete protein:protein interaction maps are now available for eukaryotes.
- The proteome is the protein version of the genome. It refers to all the proteins being expressed in a cell. The study of proteomics is becoming very important in the life sciences.
- A very powerful technique in vogue these days is the use of DNA or protein microchips. With these chips, thousands of samples of DNA or proteins can be applied and then checked for binding of biological samples.
- The binding is visualized by using fluorescently labeled molecules and scanning the chip with a computer. The pattern of fluorescent labels then gives an indication of which mRNA or proteins are being expressed in the samples.

LECTURE NOTES

This chapter covers a number of common laboratory techniques for isolating and manipulating nucleic acids. The material here serves as an excellent supplement to an associated lab course or component. One to two lectures may be devoted to this material dependent upon the students' relationship to lab work. Students who are either currently, or expect to, perform laboratory work, will find this information particularly useful. For other students, this information may still be useful for showing how basic chemical forces can be exploited in laboratory techniques. In addition, students may find these topics of particular interest as they are commonly subjects of current news, television shows, and movies.

LECTURE OUTLINE

I. Purification and detection of nucleic acids
    A. Separation techniques – gel electrophoresis
    B. Detection methods
        1. autoradiography
        2. luminescence
        3. fluorescence
II. Restriction endonucleases
    A. Palindromic cut sites
    B. "Sticky" versus blunt ends
III. Cloning
    A. Recombinant/Chimeric DNA
    B. Use of sticky ends
        1. Ligases
        2. Plasmids
        3. Bacteriophage clones
        4. Cellular clones

C. Bacteriophage and plasmid vectors
   1. Transformation
   2. Use of selectable markers
   3. Polylinker cloning sites
   4. pUC plasmids and blue/white screening as an example

IV. Genetic engineering
   A. Natural recombinant DNA
   B. Bacteria as "protein factories"
   C. Protein expression vectors
   D. Genetic engineering in eukaryotes

V. DNA libraries
   A. Creation of a library
   B. Finding an individual clone

VI. Polymerase chain reaction

VII. DNA fingerprinting
   A. Southern blots
   B. RFLPs

VIII. DNA sequencing

IX. Genomics & proteomics
   A. Microarrays
   B. Protein arrays

ANSWERS TO EXERCISES

13.1 Purification and Detection of Nucleic Acids

1. Safety, no need for special licensing, and convenience of disposal.
2. DNA is labeled with ^{32}P and run on a gel. The gel is placed next to X-ray paper, which is then developed. The radioactivity shows up as black bands on the X-ray paper. This is called an autoradiograph.
3. The DNA run on electrophoresis gels is usually cleaved with restriction enzymes to give linear pieces; thus the shape is uniform for DNA. The charge is a constant for DNA in that every nucleotide has the same charge due to the phosphate groups; thus, DNA has a uniform shape and a uniform charge-to-mass ratio, so it separates solely on size, with the shorter fragments traveling fastest through the gel.

13.2 Restriction Endonucleases

4. The use of restriction endonucleases with different specificities gives overlapping sequences that can be combined to give an overall sequence.
5. Restriction endonucleases do not hydrolyze a methylated restriction site.
6. The restriction site of the DNA of the organism that produces a restriction endonuclease is modified, usually by methylation.
7. The restriction fragments of different sizes (restriction-fragment length polymorphisms, or RFLPs) that come about as a result of different base sequences on paired chromosomes were used as genetic markers to determine the exact position of the cystic fibrosis gene on chromosome 7.

8. An endonuclease is an enzyme that cuts nucleic acid chains in the middle, as opposed to cleaving from the ends inward. The term *restriction* came from the restricted growth seen in host cells that are infected by bacteriophages when the bacteria have restriction enzymes that can cleave the viral DNA.
9. They are all palindromes (ignoring punctuation and spacing in the latter two cases), analogous to palindromic sequences of bases in DNA. Just as the five examples are distinguished by being pronounced differently, different palindromes in DNA are distinguished and acted on by different, very specific restriction endonucleases.
10. GGATCC, GAATTC, AAGCTT (remember that these are listed 5′ to 3′, so you must read the complementary strand 5′ to 3′ to see that the sequence is the same).
11. *Hae*III cuts at a sequence of four bases, cuts in the middle of the sequence, and leaves blunt ends. *Bam*HI cuts at a sequence of six bases, cuts on the second base from the 5′ end, and leaves sticky ends.
12. Sticky ends are short regions of single-stranded DNA extending from the ends of double-stranded DNA molecules. These are produced by some restriction enzymes or can be added chemically to blunt-ended double-stranded DNA. They are important because they provide a means for DNA from different sources (e.g., “foreign” gene and plasmid, both containing sticky ends) to find each other by hydrogen bonding between complementary bases. A ligase is then used to covalently link the two molecules.
13. An advantage of using *Hae*III is that it yields blunt ends. Thus, one could combine DNA cut with this enzyme with any other DNA that also had blunt ends. Enzymes exist that quickly remove the sticky overhangs from other restriction enzymes. The disadvantage is that *Hae*III is specific for a four-base sequence that is likely to occur many times in a genome, so the target DNA may also be cleaved somewhere in the middle. Also, the blunt ends make it more difficult to get specific ligation of two DNA types.

13.3 Cloning

14. A portion of exogenous DNA is introduced into a suitable vector, frequently a bacterial plasmid, and many copies of the DNA are produced when the bacteria grow. Viruses are also commonly used as vectors.
15. The most common vectors are bacterial plasmids. Viruses and cosmids can also be used, depending on the size of the foreign DNA that must be inserted.
16. The plasmid to be used as a vector needs markers both for uptake of the target DNA sequence into the plasmid and for insertion of the plasmid into host cells. Typically, a plasmid has a gene for ampicillin resistance. Only cells that have taken up a plasmid can grow on ampicillin plates. The foreign DNA is usually inserted into a second marker to select for those plasmids that took up the target DNA. This second marker may be another antibiotic-resistant gene or some other gene, such as the β-galactosidase gene.

17. The key feature of a plasmid capable of blue/white screening is the gene for the α-subunit of the enzyme β-galactosidase. These plasmids are used with a strain of *E. coli* that are deficient in the α-subunit of this enzyme. β-Galactosidase can convert a colorless sugar derivative, called X-gal, to a blue one. The site for cleavage of the plasmid by a restriction endonuclease lies within the β-galactosidase gene. Cells that have acquired a plasmid can grow on ampicillin. If the plasmid reclosed on itself without the target DNA, the colonies that took up that plasmid grow blue. Cells that have acquired the DNA insert cannot produce a blue color.
18. Restriction enzymes to cut DNA, DNA ligase to rejoin DNA, a suitable vector to carry the foreign DNA, a cell line to accept the vector, and a way of selecting for the correct transformants.
19. Since most recombinant DNA occurs with bacterial and viral vectors, a big concern is that a mutated virus or bacteria will be released that can infect other species and that may be resistant to drugs, thereby creating a new, potentially lethal disease. Precautions include frequent sterilization of cultures to make sure that they are all dead before disposal, working in laminar hoods that isolate the recombinant DNA from the outside, and care in the choice of vectors. Some vectors that are replication-deficient outside certain cell types are used so that they cannot replicate outside the lab environment.

13.4 Genetic Engineering

20. To increase disease resistance, resistance to pests, shelf life, level of nitrogen fixation (protein content), and resistance to temperature extremes.
21. Insulin, human growth hormone, tissue plasminogen activator, enterokinase, erythropoietin, and interferon.
22. The corn being grown in the field has been genetically engineered. The gene that was introduced came from the bacterium *Bacillus thuringensis.*
23. LDH 3 has the subunit composition H2M2. Each of the subunits is coded for by a separate gene, so in order to clone LDH 3, one would have to clone the gene for the M subunit and the gene for the H subunit. These would be separate cloning experiments. Each gene would be cloned into an expression cell line, and the proteins would be expressed. The individual subunits could then be combined, and they would form tetramers, some of which would be LDH 3. This could be verified by native gel electrophoresis.
24. An expression vector, such as pET 5 plasmid, has the components of any normal cloning vector (e.g., origin of replication, selectable marker, multiple cloning site), but it also has the ability to have the inserted DNA be transcribed. It has a promoter for RNA polymerase, such as T7 polymerase, and a termination sequence. These border the multiple cloning site. These vectors are used with a cell line that makes T7 RNA polymerase when induced.

25. A fusion protein is a combination of a protein coded for by an expression vector and the target gene. A common one is a histidine tag and enterokinase, which are linked to the target protein when transcribed and translated. They are used to help with the eventual purification of the target protein. The overexpressed target protein can be quickly separated from the rest of the host's proteins by purifying the fusion protein, which has characteristics that make it easy to purify.
26. The bovine growth hormone is a protein that is denatured and digested in the intestinal tract. Also, all cow's milk contains some of the hormone.
27. The DNA sequence to be inserted in the bacterial plasmid to direct the production of α-globin should be cDNA, which is a sequence complementary to the mRNA for α-globin. The cDNA can be produced on the mRNA template in a reaction catalyzed by reverse transcriptase.
28. Isolate the DNA that codes for the growth factor by means of suitable probes. Introduce the DNA into a bacterial genome. Allow the bacteria to grow and to produce human growth hormone.
29. The public is concerned about contamination with prions, which come from mammalian sources. If a mammalian protein can be expressed in large quantities in bacteria, there will be no risk of prion contamination.

13.5 DNA Libraries

30. A DNA library is a collection of cells that carry cloned pieces of the entire DNA genome of an organism. A cDNA library is made by taking the mRNA from an organism, converting it to cDNA, and cloning that for the library. In this way, the active DNA sequence is stored.
31. If a DNA library is to represent the total genome of an organism, it must contain at least one clone for each DNA sequence. This requires several hundred thousand separate clones to ensure that every sequence is represented.
32. The amount of work involved in constructing a DNA library makes it desirable to have such libraries available to the entire scientific community, thus avoiding duplication of effort.

13.6 The Polymerase Chain Reaction

33. The polymerase chain reaction depends on repeated cycles of separation of DNA strands followed by annealing of primers. The first step requires a significantly higher temperature than the second, giving rise to the requirement for strict temperature control.
34. Part of the procedure of the polymerase chain reaction requires the use of high temperatures. When a temperature-stable RNA polymerase is used, there is no need to add fresh batches of enzyme for each round of amplification. This would need to be the case, however, if the RNA polymerase could not withstand the high temperatures.
35. Good primers have similar G–C contents for the forward and reverse reactions, have minimal secondary structure possibilities with each other or with themselves, and are long enough to give sufficient specificity for the gene to be duplicated without costing too much.

36. The contaminating DNA as well as the desired DNA is amplified at each stage of the polymerase chain reaction, giving rise to an impure product.
37.
(a) The primers have very different G–C contents.
(b) The forward primer will have significant secondary structure with itself (hairpin loop due to inverted Gs and Cs on end).
(c) The forward and reverse primers will bind to each other.
38. It is a technique that allows the PCR reaction to generate time-point data that can be used to determine how much of the DNA was in the cell originally.
39. Regular PCR is designed to create large quantities of DNA, so the reaction is allowed to go to completion. With qPCR, the reaction does not go to completion as it is the time point data that are needed in order to determine the amount of the starting material.

13.7 DNA Fingerprinting

40. The polymerase chain reaction can increase the amount of a desired DNA sample by a considerable factor, making possible definite identification of DNA samples that were too small to be characterized by other means. It can be used on hair and blood samples found at the scene of a crime to establish the presence of a suspect. This method can also be used to identify remains of possible murder victims.
41. It is easier to show that two DNA samples do not match than to prove that they are identical.

13.8 Sequencing DNA

42. 5′GATGCCTACG3′
43. Two factors are involved here. First, large polymers must be cleaved into smaller, manageable fragments for sequencing. Enzymes (endoproteases) that cleave proteins, while showing some specificity, are far from absolutely specific, and messy mixtures result. On the other hand, restriction endonucleases are absolutely specific for palindromic base sequences in DNA, and “clean” cuts result, allowing easier purification. (Note that if the gene for a protein isn’t available but the mRNA is, the resulting cDNA can be made using reverse transcriptase.) A second factor is that only relatively short fragments of protein can be sequenced without additional internal cleavage. For example, the Edman degradation is limited to peptides of about 50 amino acids or fewer. With DNA, the dideoxy method coupled with polyacrylamide-gel separation can handle DNA fragments 10 to 20 times longer.
44. DNA often has introns in the gene, so knowing the DNA sequence may give the wrong answer for the final protein sequence. Also, proteins are modified posttranslationally, so there may be modifications to the protein sequence not reflected in the DNA.
45. Open-ended answer.

46. *Benefits:* A person at risk for future heart disease could be more careful with diet and exercise. Such a person might also take a drug beforehand that would help prevent the condition from developing. Doctors with access to such information would be able to make better diagnoses and to suggest quicker treatments. *Detriments:* Employment could be based on a preconceived idea of what a good genotype is. Health and life insurance could be denied to people considered to have a risky genotype. A new type of prejudice against the "genotypically challenged" could arise.

13.9 Genomics and Proteomics

47. The genome is the total DNA of a cell, containing all the genes of that organism. The proteome is the total complement of proteins.
48. A proteomic analysis has been done on the fruit fly *Drosophila melanogaster.*
49. Using robotic technology, a slide or "chip" is loaded with thousands of specific single-stranded DNA sequences. RNA is collected from samples to be tested and converted to cDNA carrying a fluorescent tag. The sample is placed over the chip and the cDNA allowed to bind. A fluorometer measures the fluorescence from the chip and indicates which DNA sequences were bound with their corresponding cDNA. This tells researchers which genes were active as only the active genes would produce RNA.
50. Yeast could be grown under the two conditions and the mRNA collected. The mRNA could then be converted to cDNA and each population could be labeled with a different color fluorescent marker. These samples could then be overlaid on a gene chip containing the yeast genome. The color of the fluorescence on the gene chip would then tell which genes were active under the two conditions.
51. Cancerous cells have altered metabolism at the genetic level. The gene expression patterns seen in patients with known types of cancer act like a fingerprint of that type of cancer. Tissue samples from patients to be diagnosed can be used to collect the RNA and convert it to cDNA. These cDNA samples are then overlaid on a gene chip of the human genome and the binding pattern analyzed through fluorescence. The pattern seen can then be compared to the patterns seen in the known cancers to aid in the diagnosis.
52. DNA microarrays have thousands of bound single-stranded DNA spots. They are used to test for the presence of the corresponding mRNA in a biological sample via cDNA produced from the mRNA. Protein arrays, on the other hand, have applied samples of very specific and pure antibodies. Biological tissue samples are placed on the protein chip. If the antigens for the specific antibodies are present, they bind to the antibodies. Another set of antibodies with fluorescent labels is then added, and the chip analyzed with a fluorometer. The patterns seen show which antigens the tissue sample had, which can be used to diagnose the patient.